Multiscale minimization of energy function for binocular stereo *

P. Compañ, R. Satorre, C. Villagrá, F. Escolano, R. Rizo
Universidad de Alicante. Dpto. Ciencia de la Computación e Inteligencia Artificial.
P.O. Box 99
03080 Alicante
SPAIN

Abstract

In this paper, we present a multi-scale stereo algorithm that minimizes an energy function at different resolutions and exploiting disparity information from previous resolution levels. The energy function is minimized using Simulated Annealing. Our scheme produces a dense disparity map. Experimental results are presented to illustrate our approach.

Keywords: stereo vision, simulated annealing, multi-scale, energy function.

1 Introduction

One of the main objectives of a vision system is to extract the depth information of the scene under consideration from its two dimensional representation. Several techniques have been developed to infer depth using only a single 2D image, but they make use of some assumptions like distant light sources, epipolar geometry, etc.

The search for the correct match of a point is called the **correspondence problem** and it is one of the central and most difficult parts of the stereo problem. For solving the correspondence problem the constraint of epipolar scanline must be assumed. In this work, this search is performed pixel by pixel but considering its neighbourhood in the epipolar line.

It is assumed that images are rectified ([1]). For a given pair of stereo images and the known orientation parameters of the cameras [2], the corresponding points are supposed to lie on the epipolar lines. Since a parallel camera alignment is used in this project, the epipolar lines are the scanlines in both images.

*Project ACTIVA: TAP98-600 by CICYT

Stochastic relaxation methods have received a lot of attention in the field of stereo vision ([3], [4]). These techniques try to obtain an optimal solution to the correspondence problem without falling in a local minimum. An example of these methods is the simulated annealing algorithm, that we will use in this paper.

On the other hand, the idea of looking at signals and analyzing them at several resolutions has received enormous attention in the field of stereo vision. **Multi-scale** or coarse-to-fine scheme, is a mode of efficiently and effectively representing data with the objective of reducing the computational complexity, [5], [6]. We will also follow a multi-scale approach in this paper.

In Section 2, we describe a new energy function to formulate the correspondence problem. Such a function incorporates information about intensity values, edges and constraints. The algorithm used for minimizing the energy function and obtaining the disparity map is described in Section 3. Section 4 explains the process of calculating images at several resolutions: the multi-scale scheme. The last section shows the experimental results and our conclusions.

2 The Energy Function

We formulate the correspondence problem as the minimization of an energy function (U^{ω}) composed by four terms: U_1^{ω}, U_2^{ω}, U_3^{ω} y U_4^{ω}, each one weighted by a control parameter γ_n. The global energy function is the combination of these terms, which incorporate information about grey intensity values and vertical edges. Two constraints are used for reducing the complexity of the search: smoothing and uniqueness.

The used notation is shown in Table 1.

L	the left image
R	the right image
$vL[i, j]$, $vR[i, j]$	the vertical line field for the images. It contains a value 1 if there is an edge between pixel (i, j) and pixel $(i, j-1)$, and 0 otherwise
$hL[i, j]$, $hR[i, j]$	the horizontal line field for the images, it contains a value 1 if there is an edge between pixel (i, j) and pixel $(i-1, j)$, and 0 otherwise
$disp$	the disparity map
δ	$\delta(a, b) = 1$ IF $a \neq b$ and 0 otherwise
ω	resolution level (scale)

Table 1: Notation for the energy function

$$U(i,j)^{\omega} = \sum_{n=1}^{4} \gamma_n U_n^{\omega}(i,j) \tag{1}$$

The term U_1^{ω} is the **grey-level matching cost** at selected pixel locations. Instead of comparing one pixel in the left image with just one pixel in the right image, we consider a window around pixel (i, j) of size $(2 * length + 1) \times (2 * width + 1)$.

$$U_1(i,j)^{\omega} = \sum_{p=-length}^{+length} \sum_{q=-width}^{+width} (L[i+p, j+q] - R[i+p, j+q+disp[i,j]])^2 \tag{2}$$

Term U_2^{ω} is the **edge-level matching cost**. We use vertical and horizontal edges information. It seems logical to suppose that if there is an edge in the left image, there should also be another edge in the right image corresponding to the first one but in a position displaced an amount of pixels determined by the disparity.

$$\begin{aligned} U_2(i,j)^{\omega} = \sum_{p=-long}^{+long} \sum_{q=-width}^{+width} \quad & (1-\delta)(vL[i+p, j+q], vR[i+p, j+q+disp[i,j]]) + \\ & (1-\delta)(hL[i+p, j+q], hR[i+p, j+q+disp[i,j]]) \end{aligned} \tag{3}$$

Term U_3^{ω} is the **smoothing constraint**: *We assume that disparity varies smoothly between edges*. This term switches off smoothing whenever a line field is encountered in the image. Clearly, the disparity value at two locations cannot be similar when there is an edge between them.

$$U_3(i,j)^{\omega} = (disp[i,j] - disp[i,j-1])^2 * (1 - vL[j,i]) \tag{4}$$

This last term incorporates the **uniqueness constraint**. If we limit ourselves to opaque objects, each point in the left image should have one and only one corresponding point in the right image. This is not true for transparent objects. It means that along any row i if we calculate the disparity at the j^{th} an the q^{th} column then according to the uniqueness constraint, $j + disp[i,j] \neq q + disp[i,q]$.

$$U_4(i,j)^{\omega} = \sum_{q=ini}^{end} \delta(j + disp[i,j], q + disp[i,q]) \tag{5}$$

3 Simulated Annealing and Global Minimization

We use a well-known stochastic relaxation method, called simulated annealing (SA), to obtain the global or nearly global optimum solution depending of the annealing schedule. There are many versions of SA algorithm, e.g. Metropolis algorithm, Creutz algorithm, Boltzman machine, Gibbs sampler, etc. In this paper, we adopt the **Metropolis algorithm** [7]. Starting with an arbitrary initial state, we visit each pixel and update the value of the disparity within a given range.

Considering the energy function defined in Section 2, the algorithm is applied iteratively. The proposed algorithm is shown in Table 2.

Step 1.	Assign initial temperature T.
Step 2.	For each pixel (i, j) in the disparity map,
	1. Change its value to any one of the range of disparity
	2. Calculate $\triangle U$
	3. if $\triangle U < 0$, accepted it if $e^{-\triangle U/T} > \xi$, where ξ is a randomly generated number from the uniform distribution over [0,1].
Step 3.	Lower the temperature by 0<k<1 such $T_{k+1} = kT_k$ and go back to Step 2 a fixed number of iterations.

Table 2: Simulated Annealing Algorithm

If the initial temperature is high enough and the annealing schedule does not cool down the system too rapidly, a global optimum solution is obtained [8]. However the theoretically initial temperature is usually so high that it needs many iterations to converge to the optimal solution. In the next section, we propose a multi-resolution structure to speed up, which initializes the correspondence problem with a solution coming from a lower resolution. Such an initialization helps to speed up convergence.

4 Multi-resolution

The original images are scaled down a power of two. The process consists of selecting the first row and deleting the next $2^n - 1$ rows, and so on. The process concerning the columns is analogous. The energy function explained in Section 2 is valid at all resolutions.

We start at the coarsest resolution, apply the SA algorithm for obtaining the disparity map at that resolution and pass it to the next finer resolution as:

$$\begin{aligned} disp^{\omega}_{2*i,2*j} &= 2 * disp^{\omega-1}_{i,j} \\ disp^{\omega}_{2*i+1,2*j+1} &= disp^{\omega}_{2*i,2*j} \end{aligned} \tag{6}$$

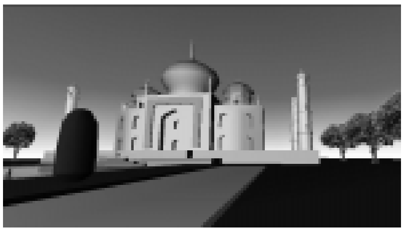

Figure 1: Original Left Image

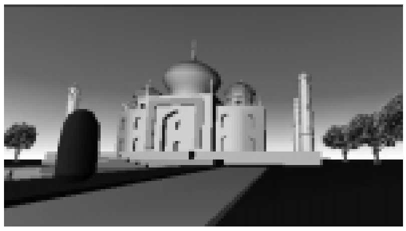

Figure 2: Left Image level 1

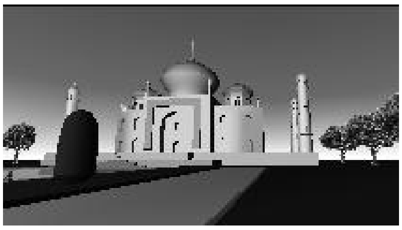

Figure 3: Left Image level 2

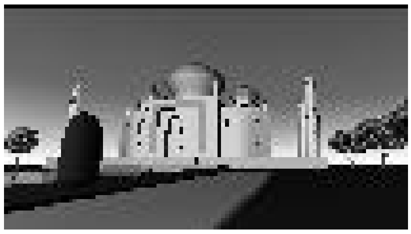

Figure 4: Left Image level 3

Some authors [9] pass the disparity map of a different manner:

$$\begin{aligned} disp^{\omega}_{2*i,2*j} &= 2 * disp^{\omega-1}_{i,j} \\ disp^{\omega}_{2*i+1,2*j+1} &= 0 \end{aligned} \tag{7}$$

We think that it is better the use of the disparity of a neighbour because of the smoothing restriction. The probability of having a disparity of a pixel similar to that of a neighbour is higher than having a disparity of zero.

The procedure finishes when the original images are processed.

5 Experimental Results and Conclusions

This section presents the experimental results for a pair of synthetic images of size 475×850 shown in Figures 5 and 6. The values assigned to the parameters are $\gamma_1 = 1$, $\gamma_2 = 120$, $\gamma_3 = 10$ and $\gamma_4 = 1$. These values were taken in order to give priority to the presence of discontinuities (edges) over smooth areas (intensities). During the scale process, we have worked with images of 237×425 (scale 2^1), 118×212 (scale 2^2) and 106×59 (scale 2^3). Figures 7 - 9 show the intermediate disparity map resulting of applying the algorithm at several resolutions. The final result is presented in Figure 10, where brighter tones indicate nearer objects.

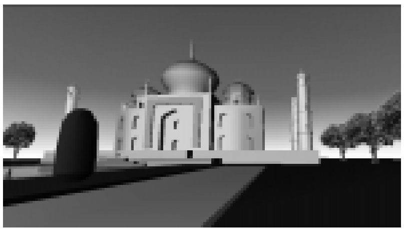

Figure 5: Left Image

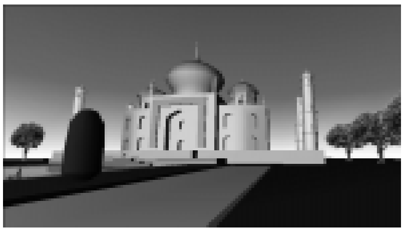

Figure 6: Right Image

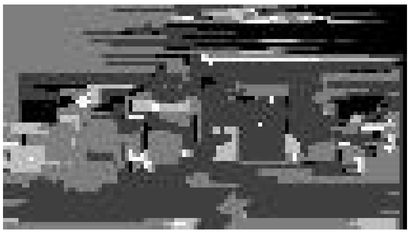

Figure 7: Result at level 3

Figure 8: Result at level 2

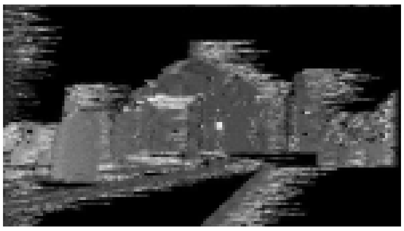

Figure 9: Result at level 1

Figure 10: Result at level 0

In previous analysis, we have worked with energy functions applied to the whole image, but we have observed that using a pixel by pixel energy function reduces the computational cost. We have performed a comparative analysis between our algorithm and a dynamic programming algorithm. As shown in Figure 11, the temporal efficiency of our algorithm increases less than the one proposed by the dynamic programming algorithm, where the image dimension is augmented steeply.

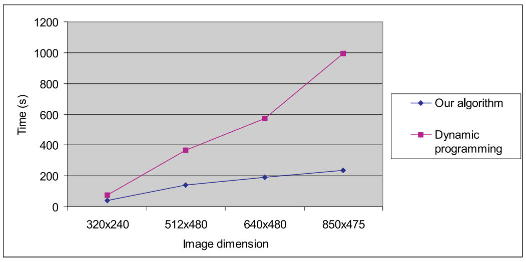

Figure 11: Comparative time analysis

We have incorporated intensity features, edges, smoothness and uniqueness for building a robust energy function.

We think that colour features will probably allow a better discrimination of pixels, so we are interested in adapting the present algorithm to a colour version. These results will be reported in the future.

References

[1] O. Faugeras. *Three-Dimensional Computer Vision. a Geometric Viewpoint*. The MIT Press, 1993.

[2] E. Trucco. *Introductory techniques for 3-D computer vision*. Prentice Hall, 1998.

[3] Chienchung Chang and Shankar Chatterjee. Multiresolution stereo - A bayesian approach. *International Conference on Pattern Recognition*, 908–912, 1990.

[4] E. Gamble, D. Geiger, T. Poggio and D. Weinshall. Integration of vision modules and labeling of surface discontinuities. *IEEE Tran. on Systems Man and Cybernetics*, 1989.

[5] S. Barnard. Stochastic stereo matching over scales. *Int. Journal of Computer Vision*, 1989.

[6] M. Lew , K. Wong and T. Huang. Multi-scale stereo matching. *International Conference on Pattern Recognition*, 620–623, 1992.

[7] N. Metropolis. Equation of state calculations by fast computing machines. *Journal Chem.*, 21: 1087–1091, 1953.

[8] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. on Pattern Anal. and Mach. Intel*, PAMI-6: 721–741, 1984.

[9] K. Sunil and U. Desai. New algorithms for 3D surface description form binocular stereo using integration. *Journal of the Franklin Institute*, 331B(5): 531–554, 1994.